Using the TeraGrid for NOAA Scientific Computing

FY 2003 NOAA HPCC NGI Project Review

Mark Govett, Dan Schaffer, Paul Hyder
(FSL)

Chris Moore, Al Hermann

(PMEL)

May 2004

Objectives and Milestones

- Determine the best grid “middleware” available
- Install and test the middleware on FSL Linux clusters
- Build a prototype coupled model where boundary condition exchanges occur across a grid
- Analyze coupled model performance across the TeraGrid
- Release version of the Scalable Modeling System that supports Grid computing
- Write Final Report

Results Summary

- Determined Globus toolkit to be best middleware
  - Used extensively throughout the US
  - Gaining corporate backing
  - Used on the TeraGrid
  - Augmented by many s/w packages that build on the toolkit
- Used Globus to construct a rudimentary grid connecting NOAA clusters
  - Two clusters at FSL and one at PMEL
  - Grid access handled with SimpleCA certificate authority

Results Summary (contd)

- Constructed prototype coupled model
- However, we were able to go further and couple Weather Research and Forecast (WRF)/Regional Ocean Modeling System (ROMS) across the grid
  - Enabled by leveraging MEAD
  - Also enabled by leveraging DoD PET Coupled Modeling Infrastructure Project that uses the Argonne Model Coupling Toolkit (MCT)
  - Also enabled by leveraging National Energy Research Scientific Computing Center (NERSC) Multi-Process Handshaking (MPH) library
  - Work nicely complemented FSL/PMEL network connection tuning effort

Results Summary (contd)

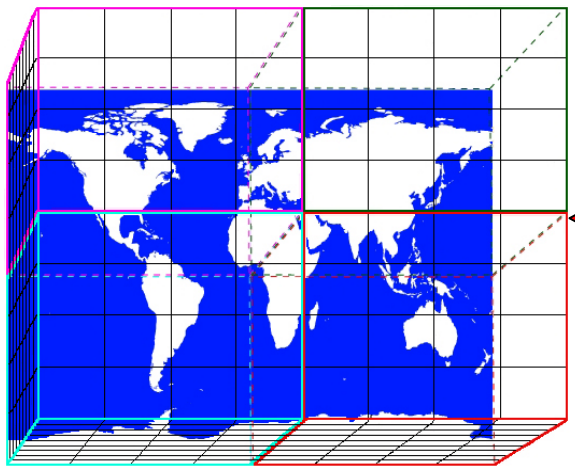
- Not able to do any coupled modeling across the TeraGrid
  - Access not granted until February 2004
  - Spent a few weeks trying to setup a simple cross-grid MPI program
  - Then the TeraGrid went down 3/30 to 5/5 due to hackers
  - Just getting going again now

Results Summary (contd)

- Released version 2.8 of SMS that supports grid computing
  - Tested against grid enabled MPICH-G2 library
  - Added support for coupled modeling

  www-ad.fsl.noaa.gov/ac/sms.html
- Completed final report including discussion of coupled model performance across the grid
  - See website www-ad.fsl.noaa.gov/ac/schaffer/ngi.html
- Paper entitled “Coupling an Oceanic/Atmospheric Model over a Geographically Distributed Computational Grid” in progress

Focus : Coupled Model



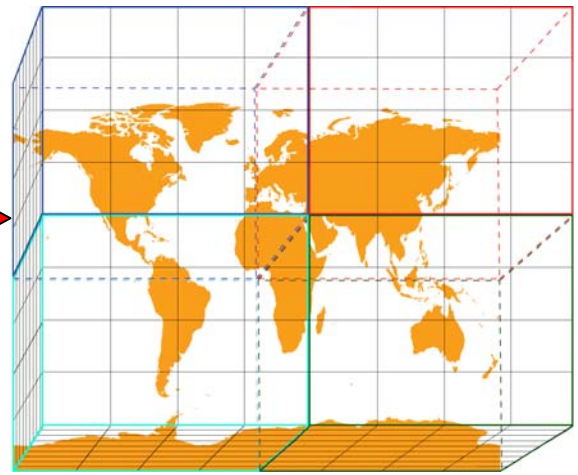
ROMS

PMEL

1200 miles

Abilene Network (100 mbit/s)

Surface boundary condition exchange



WRF

FSL

Coupled Model Focus

- Motivation
  - Coupled modeler needs 32 nodes for required time to solution but only 16 nodes available at each site AND/OR
  - Large datasets specific to each model located at each site

Coupled Model Focus

- Test case
  - WRF
    - Resolution : 161x161x30, time step = 60 seconds
    - Executed for 25 time steps for this test
  - ROMS
    - Resolution : 200x200x15, time step = 60 seconds
    - Executed 25 time steps for this test

Coupled Model Performance

- ROMS sends SST to WRF
- WRF sends to U and V Stress to ROMS
- Coupling frequency : every 5 time steps

Coupled Model Focus

- Performance
  - WRF runs on 6 processors; main model loop time 125 seconds
  - ROMS runs on 2 processors; main model loop time 52 seconds

Case	WRF Comm	ROMS Comm
Cross-grid comm (Serial, ROMS untuned)	1.53	1.15
Cross-grid comm (Serial, ROMS tuned)	1.58	0.34
Cross-grid comm (Parallel, ROMS untuned)	0.83	0.64
Cross-grid comm (Parallel, ROMS tuned)	0.78	0.51
Local comm (Parallel)	0.08	0.18

Coupled Model Focus

- Analysis
  - Tuned Serial Bandwidth ~ 60 Mbits/s
    - Compare to observed large file transfer rates (80 Mbit/s)
      - Model not run long enough to get max rate
    - Coupling frequencies can be as small as once per model day
    - Underlying communication code (MCT) has a performance bug where it unnecessarily communicates twice as much data!
    - Bandwidth may not even be an issue on TeraGrid (10 Gbit/s)
  - Latency : 0.015 seconds
    - Travel time of light between Boulder and Seattle via SF is 0.010 seconds
    - The rest is router latencies (9 router hops!!)
- Conclusion : Cross-Grid Coupled Modeling is feasible!!!

Focus : Security

- Grid work slowed by hackers and response to them
- Fundamental tension between
  - Security needs - Close off inter-machine connections
  - Grid needs - Open up inter-machine connections
    - For ease of use
    - For performance : parallel communication in a coupled model
- But a NOAA Grid will have to adapt to security concerns
  - Live with a minimum of open ports (serial communication)
  - Pay the price of VPN encryption and decryption

FY ’04 Proposal
Development of a Prototype NOAA Computational Grid

Mark Govett, Dan Schaffer,Paul Hyder
(FSL)

Chris Moore
(PMEL)

Brian Gross, Ron Bewtra
(GFDL)

Objectives

- Extend rudimentary grid to include more clusters at FSL, PMEL, GFDL
- Construct a meta-scheduler
  - Will allow NOAA-wide access to each of the HPCS's
  - Will allow priority to be given to local users
  - Meta-scheduling capability will support HPCS consolidation push

Objectives (contd)

- Execute WRF/ROMS across two medium-sized clusters at FSL and GFDL
- Execute WRF/ROMS across the TeraGrid
- Port and test GFDL CM2 coupled model on the prototype grid
- Extend the certificate authority
  - Use the NCSA MyProxy Certificate database s/w
  - Integrate certificate proxy generation with token cards (one-time passwords)

Benefits

- Capability fits in nicely with HPCS consolidation push
- Will facilitate porting of models to heterogeneous platforms
  - Helps modelers find bugs
- Will help utilize excess HPCS cycles
- Will facilitate ensemble forecasts